Pokemon Classification Problem

Final presentation

Team members: Hardo Niit, Hans Erik Atonen, Kelli Kukk, Annika Altoja

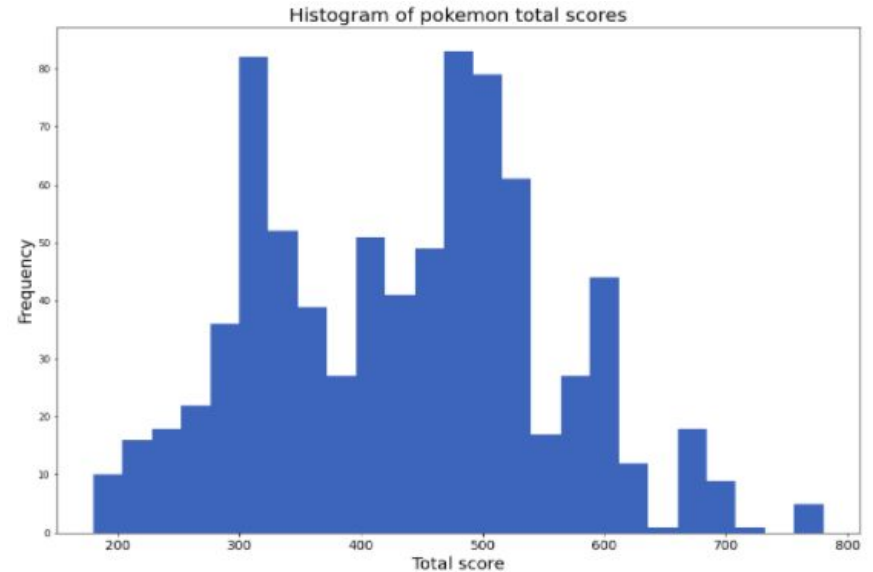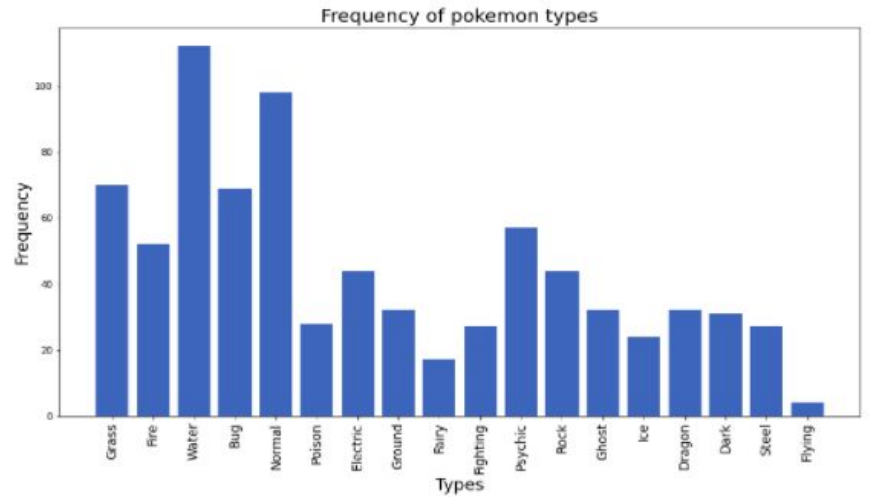Link to github: https://github.com/Hardo125/ML_project

Data

- Data comes from Kaggle
- Pokemon dataset
- 800 observation (721 unique Pokemons) 12 describing variables

What we have done?

- Statistical analysis
- Bar plots, histograms etc.
- Multiple ML models
- Ensemble models
- PCA method
- Accuracy and AUC

Frequency of pokemon types

Histogram of pokemon total scores

Results

- Used models to predict pokemon type
  Best model was RF. Accuracy under 30%
- Used models to predict legendary pokemon
  Best model was unclear. All model's had high accuracy (near 90% or above)
- PCA method to predict legendary pokemon.
  Best model was still unclear. All models retained high accuracy (90% or above)
- Found the AUC of the models predicting legendary pokemon.
  All the models are worse than they seemed. Best model seems to be RF.

Problems we faced?

- No significant problems occured…
-
- …But main problem was lack of time (other courses, homeworks, projects etc.)
-
- In addition the model's didn't turn out too great

What we learned

- We cemented all the skills we learned during the course

Thank you!